

On the relationship between empirical likelihood and empirical saddlepoint approximation for multivariate M -estimators

BY ANNA CLARA MONTI

*Centro di Specializzazione e Ricerche Economico-agrarie per il Mezzogiorno,
Via Università 96, 80055 Portici (Napoli), Italy*

AND ELVEZIO RONCHETTI

*Faculty of Economic and Social Sciences, University of Geneva, CH-1211 Geneva,
Switzerland*

SUMMARY

By comparing the expansions of the empirical log-likelihood ratio and the empirical cumulant generating function calculated at the saddlepoint, we investigate the relationship between empirical likelihood and empirical saddlepoint approximations. This leads to a nonparametric approximation of the density of a multivariate M -estimator based on the empirical likelihood and, on the other hand, it provides nonparametric confidence regions based on the empirical cumulant generating function. Some examples illustrate the use of the empirical likelihood in saddlepoint approximations and vice versa.

Some key words: Empirical bootstrap likelihood; Empirical likelihood; Influence function; M -estimator; Nonparametric confidence regions; Saddlepoint approximation; Small sample asymptotics.

1. INTRODUCTION

The purpose of this paper is to discuss the relationship between empirical likelihood techniques and empirical saddlepoint approximations for multivariate M -estimators. Owen (1988, 1990) defined an empirical version of the likelihood ratio and showed how it can be used to construct nonparametric confidence regions. In particular, he showed that, as in the parametric case, minus twice the empirical log-likelihood ratio has asymptotically a χ^2 distribution. DiCiccio, Hall & Romano (1989), DiCiccio & Romano (1989) and Hall (1990) investigated further the relationship between empirical likelihood and the parametric theory. For instance, they showed that a Bartlett correction can be applied to the empirical likelihood. Finally, Hall & La Scala (1990) give a review of the methodology and the algorithms, and Davison, Hinkley & Worton (1992) show the relationship with bootstrap likelihoods.

Saddlepoint approximations in statistics go back to Daniels (1954). These techniques provide very accurate approximations of the distribution of some statistic even in small samples. For a review, see Reid (1988), Barndorff-Nielsen & Cox (1989, Ch. 4, 6) and Field & Ronchetti (1990). The construction of these approximations typically requires knowledge of the underlying distribution F of the observations but one can make these techniques nonparametric by replacing F by the empirical distribution function. This can be viewed as a way to avoid resampling by approximating the bootstrap distribution by means of saddlepoint techniques; see Davison & Hinkley (1988), Feuerverger (1989), Young & Daniels (1990), Wang (1990) and Ronchetti & Welsh (1993).

In § 2 we summarize the results based on the empirical saddlepoint approximation. In order to show the relationship between empirical likelihood techniques and empirical saddlepoint approximations, we compare in § 3 the expansions of minus twice the empirical log-likelihood ratio $\hat{W}(t)$ and the empirical cumulant generating function $\hat{K}(t)$ for a general multivariate M -estimator. Here $\hat{K}(t) = \ln [n^{-1} \sum \exp \{\hat{\alpha}(t)' \psi(x_i, t)\}]$, where x_1, \dots, x_n are observations from a random sample, $\hat{\alpha}(t)$ is the saddlepoint which solves equation (5) below, and $\psi(x, t)$ is the vector function which defines the M -estimator. The relation is

$$n\hat{K}(t) = -\frac{1}{2}\hat{W}(t) + \frac{1}{6}n^{-1}\Gamma(u) + O(n^{-1}), \quad (1)$$

where $u = n^{1/2}(t - T_n)$, T_n is the estimator and $\Gamma(u)$ is a correction term given by (7) below and depending on the third moment of the influence function of the M -estimator. This formula allows one to translate between empirical likelihood ratios and saddlepoint densities by making a skewness correction and incurring a loss of accuracy of $O(n^{-1})$. This relationship holds generally for a statistic admitting an influence function. Moreover, the connection between the saddlepoint and the first derivative of the empirical likelihood can be related to a result by DiCiccio, Field & Fraser (1990). Finally in § 4, some examples illustrate the use of the empirical likelihood in saddlepoint approximations and vice versa.

2. THE EMPIRICAL SADDLEPOINT APPROXIMATION

Let x_1, \dots, x_n be n independent and identically distributed random variables with distribution F depending on a k -vector parameter θ belonging to some parametric space Θ . An M -estimator for θ is a statistical functional defined as a root T_n of $\sum \psi(x_i, T_n) = 0$, where the summation is over the range $i = 1, \dots, n$. Under the conditions (i)–(v) of Ronchetti & Welsh (1993), the empirical saddlepoint approximation of the density of T_n at $t = T_n + n^{-1/2}u$ is given by

$$\hat{f}_n(t) = \left(\frac{n}{2\pi}\right)^{1/2} \det \{\hat{K}''(t)\}^{-1/2} |\det \{\hat{A}(t)\}| \exp \{n\hat{K}(t)\}, \quad (2)$$

where

$$\hat{K}(t) = \ln \left[n^{-1} \sum_{i=1}^n \exp \{\hat{\alpha}(t)' \psi(x_i, t)\} \right], \quad (3)$$

$$\hat{A}(t) = \exp \{-\hat{K}(t)\} n^{-1} \sum_{i=1}^n \frac{\partial \psi(x_i, t)}{\partial t} \exp \{\hat{\alpha}(t)' \psi(x_i, t)\}, \quad (4)$$

$$\hat{K}''(t) = \exp \{-\hat{K}(t)\} n^{-1} \sum_{i=1}^n \psi(x_i, t) \psi(x_i, t)' \exp \{\hat{\alpha}(t)' \psi(x_i, t)\},$$

and $\hat{\alpha}(t)$ satisfies

$$\sum_{i=1}^n \psi(x_i, t) \exp \{\hat{\alpha}(t)' \psi(x_i, t)\} = 0. \quad (5)$$

Formula (2) can be viewed as an approximation of the bootstrap distribution which does not require resampling; see Davison & Hinkley (1988) and Ronchetti & Welsh (1993).

In order to compare the empirical saddlepoint approximation with the empirical likelihood, we consider the following expansion of $\hat{K}(t)$ (see Appendix):

$$\begin{aligned}\hat{K}(t) = & -\frac{1}{2}n^{-1}u' \hat{V}^{-1}u + \frac{1}{2}n^{-3/2}u' \hat{A}(T_n)' \hat{\Sigma}_\psi^{-1} B(u) \hat{\Sigma}_\psi^{-1} \hat{A}(T_n)u \\ & - \frac{1}{2}n^{-3/2}u' \hat{A}(T_n)' \hat{\Sigma}_\psi^{-1} c(u) - \frac{1}{6}n^{-3/2} \Gamma(u) + O(n^{-2}),\end{aligned}\quad (6)$$

where $\hat{V} = \hat{A}(T_n)^{-1} \hat{\Sigma}_\psi \{ \hat{A}(T_n)' \}^{-1}$ is the estimated asymptotic covariance matrix of T_n ,

$$\hat{\Sigma}_\psi = n^{-1} \sum_{i=1}^n \psi(x_i, T_n) \psi(x_i, T_n)', \quad \hat{A}(T_n) = n^{-1} \sum_{i=1}^n \left. \frac{\partial \psi(x_i, t)}{\partial t} \right|_{T_n},$$

$B(u)$ is a $k \times k$ matrix with elements

$$B(u)_{lj} = n^{-1} \sum_{i=1}^n \sum_{r=1}^k \left\{ \psi_l(x_i, T_n) \left. \frac{\partial \psi_j(x_i, t)}{\partial t_r} \right|_{T_n} + \psi_j(x_i, T_n) \left. \frac{\partial \psi_l(x_i, t)}{\partial t_r} \right|_{T_n} \right\} u_r,$$

$c(u)$ is a k -vector,

$$c(u) = n^{-1} \sum_{i=1}^n \sum_{j=1}^k \sum_{r=1}^k \left. \frac{\partial^2 \psi(x_i, t)}{\partial t_j \partial t_r} \right|_{T_n} u_j u_r,$$

$$\Gamma(u) = n^{-1} \sum_{i=1}^n \{u' \hat{A}(T_n)' \hat{\Sigma}_\psi^{-1} \psi(x_i, T_n)\}^3 = -n^{-1} \sum_{i=1}^n \{u' \hat{V}^{-1} \text{IF}(x_i; T_n)\}^3, \quad (7)$$

and $\text{IF}(x_i; T_n) = -\hat{A}(T_n)^{-1} \psi(x_i, T_n)$ is the influence function of T_n (Hampel, 1974). In the case of a univariate M -estimator, (6) becomes

$$\begin{aligned}\hat{K}(t) = & -\frac{1}{2}n^{-1} \frac{u^2}{\hat{V}} + n^{-3/2} \frac{u^3 \hat{A}(T_n)^2}{\hat{\Sigma}_\psi^2} n^{-1} \sum_{i=1}^n \left\{ \left. \frac{\partial \psi(x_i, t)}{\partial t} \right|_{T_n} \psi(x_i, T_n) \right\} \\ & - \frac{1}{2}n^{-3/2} \frac{u^3 \hat{A}(T_n)}{\hat{\Sigma}_\psi} n^{-1} \sum_{i=1}^n \left. \frac{\partial^2 \psi(x_i, t)}{\partial t^2} \right|_{T_n} - \frac{1}{6}n^{-3/2} \frac{u^3 \hat{A}(T_n)^3}{\hat{\Sigma}_\psi^{3/2}} \hat{\gamma}_\psi + O(n^{-2}),\end{aligned}\quad (8)$$

where

$$\hat{\gamma}_\psi = n^{-1} \sum_{i=1}^n \frac{\psi(x_i, T_n)^3}{\hat{\Sigma}_\psi^{3/2}}$$

is the third empirical cumulant of $\psi(x, t)$, given $t = T_n$. Moreover, if T_n is the univariate mean \bar{x} , then $\hat{A}(T_n) = -1$ and (8) simplifies further to

$$\hat{K}(t) = -\frac{1}{2}n^{-1} \frac{u^2}{\hat{\Sigma}} + \frac{1}{6}n^{-3/2} \frac{u^3}{\hat{\Sigma}^{3/2}} \hat{\gamma} + O(n^{-2}), \quad (9)$$

with

$$\hat{\gamma} = n^{-1} \sum_{i=1}^n \frac{(x_i - \bar{x})^3}{\hat{\Sigma}^{3/2}}, \quad \hat{\Sigma} = n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Notice that in this case $-n\hat{K}(t) = l_{\text{EB}}(t)$, the bootstrap likelihood of Davison et al. (1992, eqn (4.8)), where saddlepoint approximations are applied in place of resampling to compute a bootstrap likelihood. Therefore,

$$l_{\text{EB}}(t) = -n\hat{K}(t) = \frac{1}{2} \frac{u^2}{\hat{\Sigma}} - \frac{1}{6} n^{-1/2} \frac{u^3}{\hat{\Sigma}^{3/2}} \hat{\gamma} + O(n^{-1}). \quad (10)$$

In the notation of Davison et al. (1992), $\psi = u/\hat{\Sigma}^{1/2}$, $\kappa_{0,0} = \hat{\Sigma}$, $\kappa_{0,0,0} = \hat{\gamma}\hat{\Sigma}^{3/2}$ and our (10) corresponds to their (7.10) except that in (7.10) all signs must be reversed and the coefficient $\frac{1}{2}$ in the term of order n^{-1} must be changed to $\frac{1}{6}$.

In the next section, (6) will be compared to the expansion of the empirical log-likelihood ratio.

3. SADDLEPOINT APPROXIMATION THROUGH EMPIRICAL LIKELIHOOD

The empirical likelihood ratio is a test variable which leads to the construction of confidence regions in a nonparametric framework (Owen, 1988, 1990). The value t is included in the confidence region for θ if, according to the value of the empirical likelihood ratio, the hypothesis $H_0: \theta = t$ cannot be rejected. Formally, the empirical likelihood ratio is the ratio between the nonparametric estimate of the maximum of the likelihood function under the hypothesis H_0 and the nonparametric estimate of the maximum of the likelihood function without restrictions. Since the maximum of the nonparametric likelihood function under no restriction is n^{-n} , the empirical likelihood ratio is given by:

$$\hat{\lambda}(t) = \frac{\hat{L}(t)}{\hat{L}(T_n)} = \max_{\sum \psi(x_i, t)p_i = 0} \prod_{i=1}^n np_i,$$

where the p_i -weights satisfy

- (i) $p_i \geq 0, \quad \forall i,$
- (ii) $\sum_{i=1}^n p_i = 1,$
- (iii) $\sum_{i=1}^n \psi(x_i, t)p_i = 0.$

A Lagrangian argument leads to

$$\hat{W}(t) = -2 \ln \hat{\lambda}(t) = 2 \sum_{i=1}^n \ln \{1 + \hat{\xi}(t)' \psi(x_i, t)\},$$

where $\hat{\xi}(t)$ satisfies

$$\sum_{i=1}^n \frac{\psi(x_i, t)}{1 + \hat{\xi}(t)' \psi(x_i, t)} = 0. \quad (11)$$

The Lagrangian multiplier $\hat{\xi}(t)$ plays the same role in the construction of the empirical likelihood ratio as the saddlepoint $\hat{\alpha}(t)$ in the empirical saddlepoint approximation. In fact, for any $t \in \Theta$, the empirical likelihood ratio requires the computation of the weights p_i which centre the empirical distribution at t .

Applying the same algebra of the Appendix yields in the same notation

$$\begin{aligned} \hat{\xi}(t) = & n^{-1} \hat{\Sigma}_{\psi}^{-1} \hat{A}(T_n)u + \frac{1}{2} n^{-1} \hat{\Sigma}_{\psi}^{-1} c(u) - n^{-1} \hat{\Sigma}_{\psi}^{-1} B(u) \hat{\Sigma}_{\psi}^{-1} \hat{A}(T_n)u \\ & + n^{-2} \sum_{i=1}^n \{u' \hat{A}(T_n)' \hat{\Sigma}_{\psi}^{-1} \psi(x_i, T_n)\}^2 \hat{\Sigma}_{\psi}^{-1} \psi(x_i, T_n) + O(n^{-3/2}), \end{aligned} \quad (12)$$

and replacing (12) in the Taylor expansion of $\hat{W}(t)$ gives

$$\begin{aligned} \hat{W}(t) = & u' \hat{V}^{-1} u - n^{-1} u' \hat{A}(T_n)' \hat{\Sigma}_{\psi}^{-1} B(u) \hat{\Sigma}_{\psi}^{-1} \hat{A}(T_n)u \\ & + n^{-1} u' \hat{A}(T_n)' \hat{\Sigma}_{\psi}^{-1} c(u) + \frac{1}{2} n^{-1} \Gamma(u) + O(n^{-1}). \end{aligned} \quad (13)$$

In case of a univariate M -estimator (13) reduces to

$$\begin{aligned}\hat{W}(t) = & \frac{u^2}{\hat{V}} - 2n^{-1} \frac{u^3 \hat{A}(T_n)^2}{\hat{\Sigma}_\psi^2} n^{-1} \sum_{i=1}^n \left\{ \frac{\partial \psi(x_i, t)}{\partial t} \right\} \bigg|_{T_n} \psi(x_i, T_n) \Big\} \\ & + n^{-1} \frac{u^3 \hat{A}(T_n)}{\hat{\Sigma}_\psi} n^{-1} \sum_{i=1}^n \frac{\partial^2 \psi(x_i, t)}{\partial t^2} \bigg|_{T_n} + \frac{2}{3} n^{-1} \frac{u^3 \hat{A}(T_n)^3}{\hat{\Sigma}_\psi^{3/2}} \hat{\gamma}_\psi + O(n^{-1}),\end{aligned}$$

and, for $\psi(x, t) = x - t$,

$$\hat{W}(t) = \frac{u^2}{\hat{\Sigma}} - \frac{2}{3} n^{-1} \frac{u^3}{\hat{\Sigma}^{3/2}} \hat{\gamma} + O(n^{-1}).$$

Owen (1988) proved that $\hat{W}(t)$ is asymptotically distributed as a χ_k^2 . Thus, asymptotic confidence regions with nominal coverage probability $1 - \varepsilon$ can be constructed as

$$\{t: \hat{W}(t) \leq \chi_{k, 1-\varepsilon}^2\}, \quad (14)$$

where $\chi_{k, 1-\varepsilon}^2$ is the 100(1 - ε) percentile of the χ_k^2 distribution.

We are now able to establish the connections between the empirical likelihood and the empirical saddlepoint approximation. In fact, by a comparison of (6) and (13) we obtain formula (1),

$$n\hat{K}(t) = -\frac{1}{2}\hat{W}(t) + n^{-1}\frac{1}{6}\Gamma(u) + O(n^{-1}),$$

where $\Gamma(u)$ is defined by (7).

More generally, equation (1) holds for any multivariate statistic admitting an influence function by replacing the function ψ in \hat{K} by the influence function of the statistic.

Relation (1) can be used in two ways. First, we can obtain a nonparametric approximation of the density of T_n directly from the empirical log-likelihood ratio by replacing $n\hat{K}(t)$ by $-\frac{1}{2}\hat{W}(t) + \frac{1}{6}n^{-1}\Gamma(u)$ in the empirical saddlepoint approximation. Secondly, we can construct nonparametric confidence regions by replacing $\hat{W}(t)$ in (14) by $-2n\hat{K}(t) + \frac{1}{3}n^{-1}\Gamma(u)$. Both methods involve solving a set of equations, that is (5) and (11). In general these are of the same computational complexity. However, if one set of equations is harder to solve in a specific situation, then relation (1) allows us to use the other set as an approximation. From a conceptual viewpoint, the possibility to use the empirical likelihood to approximate the density of an estimator is attractive. This is in fact a nonparametric extension of the well known relationship between the likelihood function and the density of the maximum likelihood estimator in a parametric setup as given by Barndorff-Nielsen (1983).

Notice that the correction term in (1) depends on the skewness of $\text{IF}(x; t)$, for $t = T_n$. In the univariate case, $n^{-1}\frac{1}{6}\Gamma(u) = -u^3 \hat{V}^{-3/2} a$, where

$$a = \frac{1}{6} \sum_{i=1}^n \text{IF}(x_i; T_n)^3 / \left\{ \sum_{i=1}^n \text{IF}(x_i; T_n)^2 \right\}^{3/2}$$

is the nonparametric estimator of the acceleration constant appearing in the BC_a method of Efron (1987, eqn (7.3), p. 178). If $\text{IF}(x; T_n)$ has a symmetric distribution, $n\hat{K}(t) = -\frac{1}{2}\hat{W}(t) + O(n^{-1})$. In the case of the mean, (1) gives the correction term for the empirical bootstrap likelihood $l_{\text{EB}}(t)$ to be at distance $O(n^{-1})$ from the empirical log-likelihood ratio.

There is another point to be highlighted. If we look at (3)–(4) we notice that

$$\hat{\alpha}(t) = \{\hat{A}(t)\}^{-1} \frac{\partial \hat{K}(t)}{\partial t},$$

and replacing (1) we have

$$\hat{\alpha}(t) = -\frac{1}{2}n^{-1}\{\hat{A}(t)'\}^{-1}\frac{\partial \hat{W}(t)}{\partial t} + \frac{1}{6}n^{-3/2}\{\hat{A}(t)'\}^{-1}\frac{\partial \Gamma(u)}{\partial t} + O(n^{-3/2}).$$

Similar algebra shows that the same relationship holds between $\alpha(t)$ and $K(t)$ in the parametric saddlepoint approximation based on the true underlying distribution F as given by Field (1982). Since

$$\hat{W}(t) = -2 \ln \hat{L}(t) + 2 \ln \hat{L}(T_n),$$

we obtain

$$\hat{\alpha}(t) = n^{-1}\{\hat{A}(T_n)'\}^{-1}\frac{\partial \ln \hat{L}(t)}{\partial t} + O(n^{-1}). \quad (15)$$

This generalizes the remark on p. 94 of DiCiccio et al. (1990) that the role of the saddlepoint is being played by the first derivative of the log-likelihood function.

4. EXAMPLES

In this section we discuss four examples which illustrate the use of relationship (1).

Example 1: Duck data. We first consider the duck data taken from Larsen & Marx (1986, p. 440). It is a data set of 10 bivariate observations, where the first variable is an index of duck plumage and the second one an index of duck behaviour. Figure 1 compares the contours of empirical likelihood and normal likelihood confidence regions following Owen (1990). In Fig. 2, instead, we compare the empirical likelihood confidence regions with the saddlepoint based confidence regions. The latter are obtained by using (14),

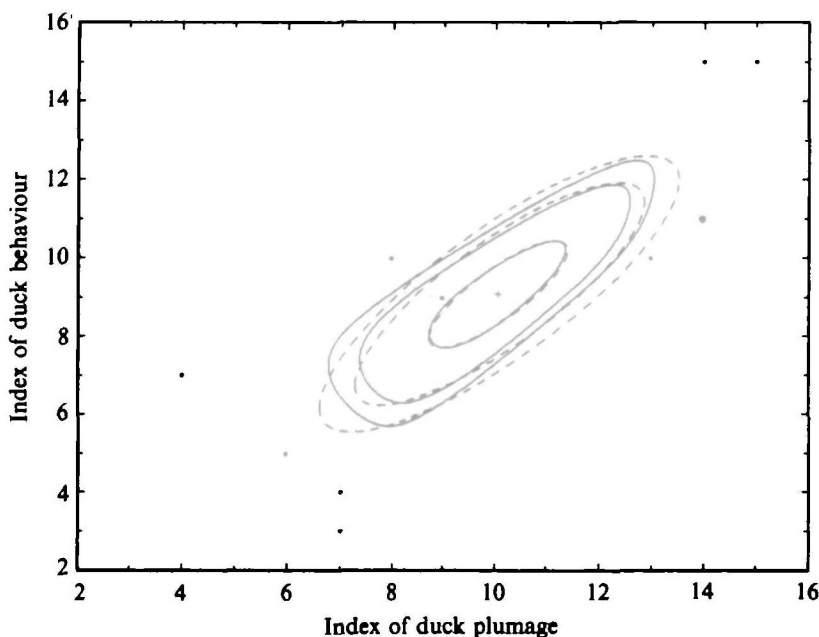


Fig. 1. 50%, 95%, 99% contours of ordinary empirical likelihood (unbroken) and normal likelihood (broken) confidence regions for duck data.

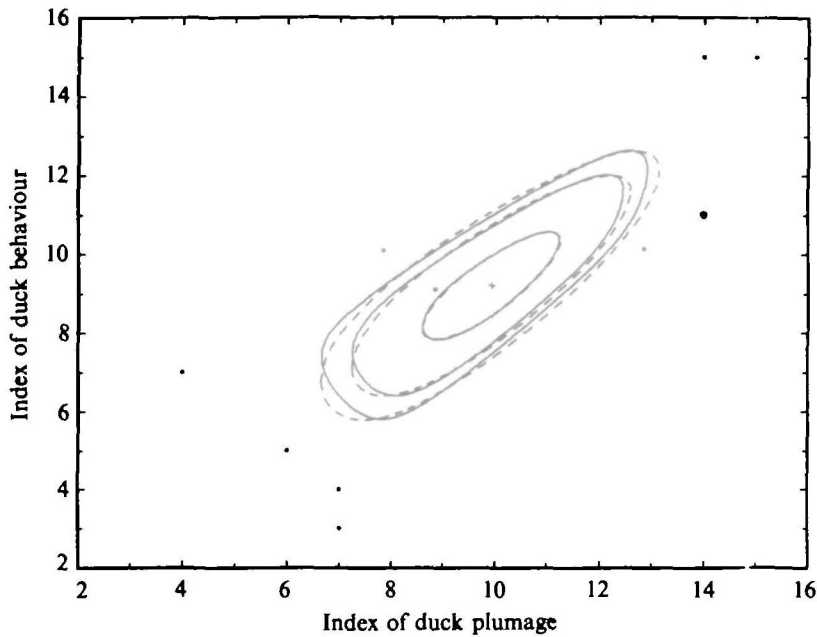


Fig. 2. 50%, 95%, 99% contours of ordinary (unbroken) and saddlepoint (broken) empirical likelihood confidence regions for duck data.

where $\hat{W}(t)$ is replaced by $-2n\hat{K}(t)$. This plot shows that even without correction term the saddlepoint based confidence regions are very close to the empirical likelihood regions.

Example 2: Bootstrap percentiles of $\bar{X} - \mu$ for the data of Davison & Hinkley (1988). Davison & Hinkley (1988, pp. 419–20) compared two approximations for the percentage points of $\bar{X} - \mu$ for a sample of size 10. They considered the empirical saddlepoint approximation and a bootstrap approximation based on 50 000 replicates. We report their numbers in the second and third column of Table 1. The fourth column gives the saddlepoint approximation obtained by replacing $n\hat{K}(t)$ by $-\frac{1}{2}\hat{W}(t) + \frac{1}{6}n^{-1}u^3\hat{\Sigma}^{-3/2}\hat{\gamma}$ as given in (1) and (9). It can be seen that the accuracy of this approximation is nearly the same as that of the original saddlepoint approximation.

Table 1. *Approximations to percentage points of $\bar{X} - \mu$, Example 2.*
Data of Davison & Hinkley (1988)

Probability	Bootstrap	Empirical saddlepoint	Empirical likelihood
0.001	-5.55	-5.52	-5.32
0.005	-4.81	-4.81	-4.59
0.010	-4.42	-4.43	-4.19
0.050	-3.34	-3.33	-3.05
0.100	-2.69	-2.69	-2.39
0.900	2.87	2.85	2.95
0.950	3.73	3.75	3.73
0.990	5.47	5.48	5.17
0.995	6.12	6.12	5.68
0.999	7.52	7.46	6.70

Bootstrap approximations based on 50 000 resamplings.

Example 3: The distribution of the Huber estimator under contaminated data. We consider an approximation of the distribution of the Huber estimator based on 10 independent and identically distributed observations from the contaminated normal distribution $0.95 \times N(0, 1) + 0.05 \times N(0, 3^2)$. The Huber estimator is an M -estimator with $\psi(x, t) = (x - t) \min(1, c/|x - t|)$ for some positive constant c . We choose here $c = 1.4$. Since the score function ψ is bounded, this estimator is robust against outlying observations and long-tailed underlying distributions. Table 2 compares the exact tail areas and those obtained by using the empirical saddlepoint and the empirical likelihood approximations. The 'exact' tail areas were obtained by a simulation on 30 000 samples. The empirical saddlepoint approximation was obtained by first computing the density $\hat{f}_n(t)$ given by (2) on a grid of points t_j at a distance 0.02 from each other. The tail areas were estimated by

$$P(T_n \geq t) = \sum_{t_j > t} \hat{f}_n(t_j) / \left\{ \sum_{t_j = -\infty}^{+\infty} \hat{f}_n(t_j) \right\}.$$

The numbers given in the third and fourth columns of Table 2 are averages over 200 samples. The empirical likelihood approximation is computed as the empirical saddlepoint approximation except that $n\hat{K}(t)$ is replaced by $-\frac{1}{2}\hat{W}(t)$. Table 2 shows that the empirical saddlepoint approximation is more accurate for tail areas between 1% and 4% while the empirical likelihood approximation outperforms the former in the range 4%–10%.

Table 2. Tail areas of the Huber estimator, Example 3
 $c = 1.4$, $X_i \sim 0.95 \times N(0, 1) + 0.05 \times N(0, 3^2)$

t	Exact	Empirical saddlepoint	Empirical likelihood
0.44	0.102	0.078	0.099
0.52	0.069	0.051	0.071
0.58	0.049	0.037	0.055
0.64	0.034	0.027	0.042
0.70	0.023	0.019	0.033
0.76	0.016	0.014	0.025
0.82	0.010	0.010	0.019

Example 4: Gesell adaptive score data. The data set is taken from Mickey, Dunn & Clark (1967) and Rousseeuw & Leroy (1987, p. 46). The Gesell adaptive score for 21 children is given with the age, in months, at which a child utters its first word. Mickey et al. (1967) found that observation 19 is an outlier, so a robust analysis is appropriate. We first centred the x 's and y 's with respect to their medians (Table 3). Then we performed a robust regression of y versus x by estimating the slope by a Huber estimator with $c = 1.4$. Figure 3 compares the empirical log-likelihood ratio $-\frac{1}{2}\hat{W}(t)$ as given by Owen (1991) with $n\hat{K}(t)$. Again the two curves are very close.

Table 3. Gesell adaptive score data (centred), Example 4

x	4	15	-1	-2	4	9	7	0	-3	9	-4
y	0	-24	-12	-4	7	-8	-2	5	9	-1	18
x	-2	-1	0	0	-1	1	31	6	0	-1	
y	1	-12	-11	7	5	10	-38	26	-9	5	

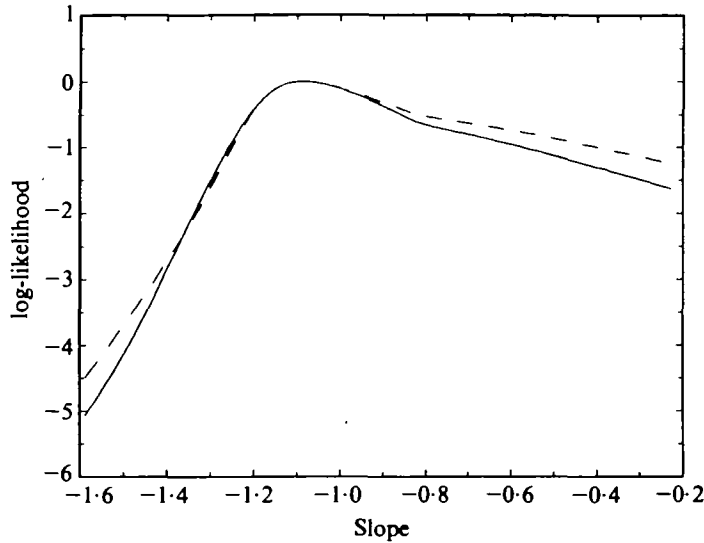


Fig. 3. Approximation of the log-likelihood of the regression coefficient in the data set of Rousseeuw & Leroy (1987) through $-\frac{1}{2}\hat{W}(t)$ (unbroken) and $n\hat{K}(t)$ (broken).

ACKNOWLEDGEMENTS

The research of the first author was partially supported by C.N.R. The authors are grateful to the editor and the referees for their comments which led to an improved exposition.

APPENDIX

Derivation of the expansion (6)

First consider the saddlepoint $\hat{\alpha}(t)$. Applying the implicit function theorem in (5) yields $\hat{\alpha}(t) = O(n^{-1})$. Thus we have

$$n^{-1} \sum_{i=1}^n \psi(x_i, t) \{1 + \hat{\alpha}(t)' \psi(x_i, t)\} = O(n^{-1}).$$

Writing $\hat{\alpha}(t) = \hat{\alpha}_1(t) + \hat{\alpha}_2(t)$, we get $\hat{\alpha}_1(t) = -n^{-1} \hat{\Sigma}_\psi^{-1} \hat{A}(T_n) u$ and $\hat{\alpha}_2(t) = O(n^{-1})$. Replacing $\hat{\alpha}(t) = \hat{\alpha}_1(t) + \hat{\alpha}_2(t)$ in (5) gives

$$n^{-1} \sum_{i=1}^n \psi(x_i, t) [1 + \{\hat{\alpha}_1(t) + \hat{\alpha}_2(t)\}' \psi(x_i, t) + \frac{1}{2} \{\hat{\alpha}_1(t)' \psi(x_i, t)\}^2] = O(n^{-3/2}).$$

Thus

$$\begin{aligned} \hat{\alpha}(t) &= \hat{\alpha}_1(t) + \hat{\alpha}_2(t) \\ &= -n^{-1} \hat{\Sigma}_\psi^{-1} \hat{A}(T_n) u - \frac{1}{2} n^{-1} \left[n^{-1} \sum_{i=1}^n \{u' \hat{A}(T_n)' \hat{\Sigma}_\psi^{-1} \psi(x_i, T_n)\}^2 \hat{\Sigma}_\psi^{-1} \psi(x_i, T_n) \right] \\ &\quad + n^{-1} \hat{\Sigma}_\psi^{-1} B(u) \hat{\Sigma}_\psi^{-1} \hat{A}(T_n) u - \frac{1}{2} n^{-1} \hat{\Sigma}_\psi^{-1} c(u) + O(n^{-3/2}). \end{aligned}$$

Substituting the expansion for $\hat{\alpha}(t)$ in (2) and expanding $\hat{K}(t)$, we obtain

$$\begin{aligned} \hat{K}(t) &= \ln \left(n^{-1} \sum_{i=1}^n [1 + \hat{\alpha}(t)' \psi(x_i, t) + \frac{1}{2} \{\hat{\alpha}(t)' \psi(x_i, t)\}^2 + \frac{1}{6} \{\hat{\alpha}(t)' \psi(x_i, t)\}^3] \right) + O(n^{-2}) \\ &= \hat{\alpha}(t)' n^{-1} \sum_{i=1}^n \psi(x_i, t) + \frac{1}{2} \hat{\alpha}(t)' n^{-1} \sum_{i=1}^n \psi(x_i, t) \psi(x_i, t)' \hat{\alpha}(t) + \frac{1}{6n} \sum_{i=1}^n \{\hat{\alpha}(t)' \psi(x_i, t)\}^3 + O(n^{-2}) \\ &= -\frac{1}{2} n^{-1} u' \hat{V}^{-1} u + \frac{1}{2} n^{-3/2} u' \hat{A}(T_n)' \hat{\Sigma}_\psi^{-1} B(u) \hat{\Sigma}_\psi^{-1} \hat{A}(T_n) u - \frac{1}{2} n^{-3/2} u' \hat{A}(T_n)' \hat{\Sigma}_\psi^{-1} c(u) - \frac{1}{6} n^{-3/2} \Gamma(u) \\ &\quad + O(n^{-2}). \end{aligned}$$

REFERENCES

- BARNDORFF-NIELSEN, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343-65.
- BARNDORFF-NIELSEN, O. E. & COX, D. R. (1989). *Asymptotic Techniques for Use in Statistics*. London: Chapman and Hall.
- DANIELS, H. E. (1954). Saddlepoint approximation in statistics. *Ann. Math. Statist.* **25**, 631-50.
- DAVISON, A. C. & HINKLEY, D. V. (1988). Saddlepoint approximations in resampling methods. *Biometrika* **75**, 417-31.
- DAVISON, A. C., HINKLEY, D. V. & WORTON, B. J. (1992). Bootstrap likelihoods. *Biometrika* **79**, 113-30.
- DI CICCIO, T. J., FIELD, C. A. & FRASER, D. A. S. (1990). Approximations of marginal tail probabilities and inference for scalar parameters. *Biometrika* **77**, 77-95.
- DI CICCIO, T. J., HALL, P. & ROMANO, J. P. (1989). Comparison of parametric and empirical likelihood functions. *Biometrika* **76**, 465-76.
- DI CICCIO, T. J. & ROMANO, J. P. (1989). On adjustments to the signed root of the empirical likelihood ratio statistics. *Biometrika* **76**, 447-56.
- EFRON, B. (1987). Better bootstrap confidence intervals (with discussion). *J. Am. Statist. Assoc.* **82**, 171-200.
- FEUERVERGER, A. (1989). On the empirical saddlepoint approximation. *Biometrika* **76**, 457-64.
- FIELD, C. A. (1982). Small sample asymptotic expansions for multivariate M -estimates. *Ann. Statist.* **10**, 672-89.
- FIELD, C. A. & RONCHETTI, E. (1990). *Small Sample Asymptotics*. Hayward, CA: Inst. Math. Statist. Monograph Series.
- HALL, P. (1990). Pseudo likelihood theory for empirical likelihood. *Ann. Statist.* **18**, 121-40.
- HALL, P. & LA SCALA, B. (1990). Methodology and algorithms of empirical likelihood. *Int. Statist. Rev.* **58**, 109-27.
- HAMPEL, R. R. (1974). The influence curve and its role in robust estimation. *J. Am. Statist. Assoc.* **69**, 383-93.
- LARSEN, R. J. & MARX, M. L. (1986). *An Introduction to Mathematical Statistics and its Applications*. Englewood Cliffs, NJ: Prentice-Hall.
- MICKEY, M. R., DUNN, O. J. & CLARK, V. (1967). Note on the use of stepwise regression in detecting outliers. *Comp. Biomed. Res.* **1**, 105-111.
- OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237-49.
- OWEN, A. B. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90-120.
- OWEN, A. B. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19**, 1725-47.
- REID, N. (1988). Saddlepoint methods and statistical inference. *Statist. Sci.* **3**, 213-38.
- RONCHETTI, E. & WELSH, A. H. (1993). Empirical saddlepoint approximations for multivariate M -estimators. *J. R. Statist. Soc. B*. To appear.
- ROUSSEEUW, P. & LEROY, A. (1987). *Robust Regression and Outlier Detection*. New York: Wiley.
- WANG, S. (1990). Saddlepoint approximations in resampling analysis. *Ann. Inst. Statist. Math.* **42**, 115-31.
- YOUNG, G. A. & DANIELS, H. E. (1990). Bootstrap bias. *Biometrika* **77**, 179-85.

[Received April 1992. Revised September 1992]